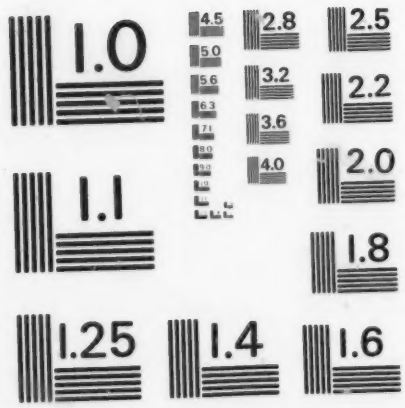
1 OF 1
ED
024528



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DOCUMENT RESUME

ED 024 528 RE 001 418

By- Rosenshine, Barak

New Correlates of Readability and Listenability.

Pub Date Apr 68

Note- 17p.; Paper presented at International Reading Association conference, Boston, Mass., April 24-27, 1968.

EDRS Price MF-$0.25 HC-$0.95

Descriptors- Evaluation Criteria, *Readability, Reading Comprehension, Reading Difficulty, *Reading Level, *Reading Material Selection, Textbook Selection

Horizontal readability, the analysis of essentially similar passages through classification of words and phrases according to their cognitive similarity, is discussed in relation to its usefulness in differentiating among materials designed for the same reading level. Three studies of horizontal readability in which passages were rated for high and low comprehension are described. Research findings indicated five variables of reading difficulty. (1) Vagueness, which resulted from indeterminate qualifiers and probability words, lowered comprehension. (2) Explaining links such as prepositions and conjunctions, which indicated that the cause, result, or means of an event or idea was being presented, raised comprehension. (3) The frequent use of examples produced greater understanding. (4) A rule-example-rule pattern of explanation was more effective than either inductive or deductive explanation. (5) The elimination of irrelevant sentences facilitated comprehension. It is noted that experimental research will be necessary to further clarify these findings, but it is recommended that the findings be implemented now for textbook evaluation. References are listed. (BS)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 024523

NEW CORRELATES OF READABILITY AND LISTENABILITY

Barak Rosenshine

Temple University

Paper delivered in session on
Research Reports: Readability

Thirteenth Annual Convention of the
International Reading Association

Boston, Massachusetts
April, 1968

RE 001 418

NEW CORRELATES OF READABILITY AND LISTENABILITY [1]

There are three steps in investigating the correlates of reading difficulty. First, investigators order a number of reading materials according to a criterion of difficulty, usually comprehension; second, they analyze the materials for internal, linguistic factors which predict variation in difficulty; and third, they combine the factors which are most predictive of difficulty into a multiple regression formula. Research by this paradigm has been fruitful, and the findings have been remarkably consistent. Almost without exception the studies have shown that difficult reading material contains longer words and longer sentences. Recent studies which have used refined measurement techniques enable us to combine a variety of additional linguistic variables into prediction equations which correlate .9 or better with the difficulty of the passages (1) (2).

In this type of research, the passages are usually short - ranging from 150 to 300 words - differ widely in content, and represent a range of difficulty of eight years or more. Traditional readability research might be labeled vertical studies because the difficulty of the passages ranges from high to low. Coleman (2) has pointed out that the predictor variables developed from such vertical studies are very useful in distinguishing among passages that range widely, such as high school and

---

1. The primary intellectual debts in this report are owed to professors N. L. Gage (Stanford University), John Bormuth and Carl Rinne (both of the Un iversity of Chicago), E. B. Coleman (University of Texas at El Paso) and my wife, Barbara.

grammar school prose.

But a more difficult problem is the development of measures which will distinguish between the effectiveness of essentially similar passages. For example, if five 5th grade American History texts are roughly equal in letters per word and words per sentence, can the teacher assume that the texts will be equally comprehensible? If not, then what measures will identify the effective and ineffective texts? The development of measures which distinguish between similar material might be termed studies of horizontal readability. This is a report of three studies of horizontal readability.

In the first study, Peterson (13) took two 950-word passages from social studies textbooks - one on feudalism and the other on imperialism - and rewrote each of the original passages into an "organized version" and a "human interest" version. The subjects were seventh-grade students.

A second researcher, Ray Funkhouser, at the Institute for Communication Research at Stanford University (7) (8) studied the problems in communicating science material to non-scientists. The experimental variable in this study was a set of 10 different eight-page articles on enzymology written for an audience of non-specialists. Nine of the articles were written by professional science writers, and the tenth was written by a biochemist. The subjects were first- and second-year college students who were not majoring in science.

In a third study, (14) (15), 40 twelfth-grade social studies teachers gave two 15-minute lectures to their students on successive days: first on Yugoslavia and then on Thailand. All lectures were prepared from identical material: 2500-word articles taken from the Atlantic magazine.

There is some question about including a study of lecturing, but the research summaries by Klare (11) and Travers (18) indicate that readability and listenability are highly correlated.

In each of the three studies, all students took a common text based on the main ideas of the original material. All three investigators adjusted the students' test scores for a measure of student knowledge or aptitude. In addition, Funkhouser (7) (8) and Rosenshine (14) adjusted the students' posttest scores for the relevance of the material in the article they read or the lecture they heard to the items on the common test.

In all three studies, although the passages contained similar material and were intended for the same audience, there were significant differences in the effectiveness of these passages as measured by the students' adjusted comprehension scores. Peterson ( ) found, as she predicted, that the students who read the organized version or the human interest version had significantly higher adjusted test scores on a common test than those who read the original passages on feudalism and imperialism. The 10 articles on enzemology, the lectures on Yugoslavia, and the lectures on Thailand all differed significantly in their effectiveness.

Results

What variables accounted for these differences? First, some negative findings. In all of these studies, the traditional readability variables and the traditional readability formulas did not discriminate between the high-comprehension producing passages and the low-comprehension producing passages. In all three studies, there were no significant differences between the passages on measures of word length, word

difficulty, and sentence length. Funkhouser (8) found that three readability measures such as the Flesch (5) and the Dale-Chall (3), and three versions of the "cloze score" (17) did have moderate and consistent correlations with the adjusted measure of comprehension, but none of the correlations was statistically significant. In addition, Peterson's materials did not differ in the number of personal pronouns or the number of personal sentences. Funkhouser found no significant differences in words per paragraph, type-token ratio, percent of lines of analogy, percent of lines of definition, and percent of lines of non-science material. In the lectures on Yugoslavia and Thailand (14)there were no significant differences in the length and structure of independent clause units, frequency and proportion of prepositional phrases, and in the use of personal reference pronouns, passive verbs, and awkward and fragmented sentences.

There were, however, five promising variables which emerged from this research. They are: vagueness, explaining links, frequency of examples, the rule-example-rule pattern, and something which might be labeled irrelevancy.

1. Vagueness. Page (12), Hiller, and others (9) have developed computer programs to count the frequencies of certain stylistic elements in essays. They have found that the "essay grades" developed from the computer count of these stylistic elements correlated significantly with the grades assigned to the same essays by humans (9) (12). One of the categories which Hiller developed for the analysis of essays was labeled vagueness and defined as a writing style characterized by an excessive proportion of qualification, haziness, and ambiguity.

Hiller, et al. (10) expanded the list of words and phrases taken to indicate vagueness and used a computer to count the proportion of vagueness words and phrases in 32 of the Yugoslavia lectures and 23 of the Thailand lectures. Hiller, et al. found that the proportion of words classified in the subcategory indeterminate qualifiers and the proportion of words classified as probability had significant negative correlations with the difficulty of both the Yugoslavia and the Thailand lectures. Indeterminate qualifiers are words such as "rather," "very," "any number of," "more or less," "little," "few," "some," "pretty much," and "quite a bit." Probability words include "could be," "might," "possibly," "sometimes," "more often than not," "may," "usually," "liklihood," and "most of the time." The high-scoring lectures, then, had fewer indeterminate qualifiers and probability words.

Hiller's findings indicate that although the use of short words usually correlates positively with reading ease, there are some pretty short words which, more often than not, might possibly detract very much from readability, more or less.

2. Explaining Links. In my analysis of the Yugoslavia and Thailand lectures (14) (15), I assessed the frequency of explanation by counting explaining links, that is, propositions and conjunctions which indicate that the cause, result, or means of an event or idea is being presented. Explaining links are words and phrases such as "because," "in order to," "if...then," "therefore," "consequently," and "by means of," as well as specified instances of words such as "since," "by," "through," and "so." The high-scoring lectures on Yugoslavia and on Thailand used more of these explaining links in each of three units of measure: per lecture, per minute, and per 100 words.

The identification of explaining links may be one step in developing a measure of the connectiveness of material. Words such as these explaining links may function to link phrases either within or between sentences so that a phrase or clause containing an explaining link elaborates and expands upon another phrase or sentence. This special linkage may be illustrated by the following three sentences which are almost identical:

(1) The Chinese dominate Bangkok's economy, and they are a threat.

(2) The Chinese dominate Bangkok's economy, but they are a threat.

(3) The Chinese dominate Bangkok's economy; therefore, they are a threat.

The third sentence may be the easiest to comprehend because it contains the explaining link "therefore" instead of a conjunction such as "and" or "but." Different types of explaining links also seem to be interchangeable, as in the following three examples:

(1) The Chinese dominate Bangkok's economy; therefore, they are a threat. (Statement of consequence)

(2) The Chinese are a threat because they dominate Bangkok's economy. (Statement of cause)

(3) By dominating Bangkok's economy, the Chinese are a threat. (Statement of means)

It should be noted that the explaining links which were counted in this study were only a convenience for identifying "explaining sentences." There is no claim that the words selected as explaining links represent all the words which could be selected. One next step will be to investigate this category more closely, eliminating words which are not true explaining links and determining whether certain nouns and verbs can be included within this category.

I also counted the number of explaining links in the passages developed by Peterson and found that the frequency of explaining links did discriminate between the high-scoring and low-scoring passages on imperialism, but that the differences were not significant for the passages on feudalism.

3. Examples. Funkhouser (7) found that for the material on enzymology, the proportion of lines giving examples was a significant positive correlate of effectiveness. In the lectures on Yugoslavia and Thailand the number of examples was not significant. Such a contrast makes sense because enzymology appears to be a more difficult topic to the reader than the political and economic affairs of Yugoslavia, and because the science articles were rated by the Flesch reading ease formula as more difficult than the social studies lectures. These results appear to indicate that the frequency of examples becomes more important as the conceptual difficulty of the material increases.

4. Rule and Example Pattern. Although the high-scoring lectures on Yugoslavia and Thailand did not contain more examples or sections of examples, they differed from the low-scoring lectures in the pattern of examples. The high-scoring lectures used a summarizing rule twice, both before and after a series of examples (15).

For example, when the high-scoring lecturers were discussing Yugoslavia's problems with inflation, they would begin with a general statement such as "Tito is attempting to deal with the problems of inflation," or "They are attempting to curb inflation." They would follow this general statement with a number of examples and close by re-stating the general statement using sentences such as "So you can see that they have a problem with inflation." Some high-scoring lecturers restated

the principle indirectly by beginning the next sentence with "In addition to the problem of inflation, Yugoslavia also..." In contrast, the low-scoring teachers used only one summary statement, usually before the series of examples.

These results indicate that a pattern which presents a structuring statement first, follows it with details, and concludes with a structuring statement is more effective than either an inductive or a deductive pattern of explanation. An extension of this idea might be that some paragraphs would be more effective if they began and ended with a topic sentence.

5. Irrelevancy. Although Funkhouser found that increased redundancy of examples is a positive correlate of effectiveness, his results also suggest that not all redundancy is useful because in his study, the number of lines relevant to each test item had a negative correlation with effectiveness. That is, the high-scoring articles had fewer lines related to an item on the test. Although this is surprising, it is not an isolated finding. Desirato, et al. (4) reduced the length of lectures, and Fletcher (6) reduced the length of film commentary by eliminating digressions and irrelevancy. In both cases the reduction in material resulted in significantly increased comprehension as measured by test scores.

The existence of irrelevancy will complicate future research in this area because irrelevancy will not be identified using the current readability formulas. In all three studies the number of words per sentence was not a significant correlate. So irrelevancy expresses itself by extra sentences, not by extra long sentences. Irrelevant material may also contain short words, an abundance of explaining links, and even

paragraphs which use the rule-example-rule pattern.

Inplications

There are two general conclusions which can be drawn from these three horizontal or limited-range studies. The first is that when relatively long passages deal with similar material and are intended for the same audience, the passages still differ in their difficulty or comprehensibility. However, in these cases the readability formulas are not particularly valid measures for distinguishing between the effective and ineffective passages.

The second conclusion is that the measures which have been related to the effectiveness of similar passages were developed by focusing upon the cognitive function of key words and phrases. The words and phrases in each of the significant findings are not linguistically or structurally similar. For example, the words classified under indeterminate qualifiers include such structurally different words as "few," "any great extent," and "more or less." Yet, these words share the cognitive function of being vague qualifiers. The words classified as explaining links include "because," "therefore," "by means of," and "in order to." Although these words are structurally dissimilar, they do have a cognitive similarity: they all introduce a clause or phrase which states a means, reason, or consequence for the idea expressed in another clause. The results are the same in the case of examples. We would be hard put to find a structural difference between a sentence giving an example, and a sentence which introduces or summarizes the topic. These results suggest that in the study of similar passages, attempts to discriminate between effective and ineffective presentations by counting only different parts

of speech may have limited promise. More significant results may be obtained by also focusing upon the cognitive function of words and phrases.

The suggestion that we consider the cognitive as well as the linguistic function of words and phrases might also be developed from the research of Coleman (2) and Bormuth. In Coleman's research, adverbs of time and location are classified separately from other adverbs. Coleman found that such a distinction has research merit. But Coleman's categories require separate classifications for the adverb "now" and for the prepositional phrase "at the present time;" separate classifications for "usually" and for "most of the time." Bormuth (personal communication) has suggested that all four examples be classified together as time adverbials. If the proportion of time adverbials in a communication is a significant correlate of reading ease, then we may obtain better research results if we classify all words dealing with time and location together, regardless of their structural use. For example, if the words "yesterday" and "today" receive special attention when they are used as adverbs, perhaps these words should receive the same attention when they are used as nouns.

The cognitive approach to the study of reading difficulty suggests not only combinations, but also new divisions of structural categories. In Coleman's report, he cites the following eight words as examples of the subclass predeterminers: each, all, both, half, any, some, most, and few. The first four words are specific, the last four words - any, some, most, and few - were among the words selected by Hiller as indeterminate qualifiers. In future studies, spearate classifications of predeterminers into specific and indeterminate may yield productive

results. But for such research we will have to use passages longer than 200 words.

Conclusions

This new research, horizontal studies of readability, consists of the analysis of essentially similar passages and focuses upon classifying words and phrases according to their cognitive similarity. This research has produced promising potential correlates of reading difficulty, such as vagueness words, explaining links, redundancy of examples, and the rule-example-rule pattern. Experimental research will be necessary to clarify these findings. This experimental research could involve inserting and deleting explaining links and vagueness words from selected passages.

It is too early to make definitive recommendations at this point, but when teachers have to choose among similar texts with relatively similar readability levels, the teacher might be aided in his choice by counting the proportion of vagueness words, rule-example-rule pattern, and explaining links in these texts.

Future research in this area will be far from simple. In both experimental and correlational studies some of these results will fail to replicate, and new variables will emerge that will be even more bewildering. There will doubtless be complex interactions among cognitive and linguistic variables, relationships with effectiveness which are not linear, and interactions and correlations which change as the material becomes more complex. There is a need for more horizontal studies such as these.

References

1. Bormuth, J. R. Readability: A new approach. Reading Research Quarterly, 1966, 1, 79-132.

2. Coleman, E. B. Developing a Technology of Written Instruction: Some Determiners of the Complexity of Prose. In E. Z. Rothkopf (Ed.) Verbal Learning Research and a Technology of Written Instruction. Rand McNally, in press.

3. Dale, E. and Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20

4. Desiderato, O. L., Kanner, J. H., and Runyon, R. P. Procedures for improving television instruction. Audio-Visual Communication Review, 1956, 4, 57-63.

5. Flesch, R. F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

6. Fletcher, R. M. Profile analysis and its effect on learning when used to shorten recorded film commentaries. Technical Report SDC 269-7-55. Port Washington, Long Island, N.Y. U.S. Naval Special Devices Center, 1955.

7. Funkhouser, G. R. Communicating science to non-scientists. Paper presented at the meeting of the American Association for Public Opinion Research, Lake George, New York, 1967.

8. Funkhouser, G. R. Readability in science writing. Paper presented at the meeting of the Association for Education in Journalism, Boulder, Colorado, 1967.

9. Hiller, J. H. and Marcotte, D. R. A computer search for the traits of opinionation, vagueness, and specificity-distinctiveness in student essays. Paper presented at the meeting of the American Psychological Association, Division 5, Washington, 1967.

10. Hiller, J. H., Fisher, G., and Kaess, W. A computer investigation of characteristics of teacher lecturing behavior. Paper presented at the meeting of the American Educational Research Association, Chicago, February, 1968.

11. Klare, G. R. The measurement of readability. Ames, Iowa: Iowa State University Press, 1963.

12. Page, E. B. The imminence of grading essays by computer. Phi Delta Kappon, 1966, pp. 238-243.

13. Peterson, E. M. Aspects of readability in the social studies. Bureau of Publications, Teachers College, Columbia University, New York: 1954.

14. Rosenshine, B. Objectively measured behavioral predictors of effectiveness in explaining. Paper presented at the meeting of the American Educational Research Association, Chicago, 1968.

15. Rosenshine, B. Behavioral predictors of effectiveness in explaining social studies material. Unpublished doctoral dissertation, Stanford University, 1968.

16. Strunk, W., Jr. and White, E. B. The elements of style. New York: Macmillian, 1965.

17. Taylor, W. "Cloze procedure:" A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

18. Travers, R. M. W. et al. Research and theory related to audiovisual information transmission. Interim Report, U.S. Department of Health, Education, and Welfare; Office of Education Contract No. 3-20-003. Salt Lake City, Utah: University of Utah, Bureau of Educational Research, 1964.

Additional Examples of the Use of Explaining Links

The four examples presented below are additional, and perhaps more dramatic, illustrations of the use of explaining links. Each example is the material which a different teacher used to cover one of the given criterion questions in the Thailand lecture. The original question, which all teachers were given was:

> Chinese citizens of Thailand, deemed by many as being prone to Communist subversion,
>
> a) dominate the country's commerce.
> b) are a small minority.
> c) are many in number but have little government influence.
> d) make up most of the lower-economic population.
>
> The correct response is a.

The four excerpts below are taken from two high-scoring lectures and two low-scoring lectures. The explaining links are underlined. The purpose of the four examples is to show the stylistic differences between the high and low lectures. All four examples cover the criterion question; in fact, all four almost quote the correct response. But within their presentation, the two high-scoring lecturers are using explaining links whereas the two low-scoring lecturers are not.

Teacher 1T

> They give an example of one of the problems that Thailand faces. They had a recent offering of 25 scholarships to anyone in the country. The winners, 20 out of 25, 23 were Chinese and 2 were native Thais.

Now this shows one thing, that the Chinese, even though they are a minority, dominate the country's business and commerce. And this is also true, I think, generally throughout Southeast Asia. Of Bangkok's population of two and a half million, about half are Chinese. That's just the capital, not the country. These Chinese, as you might imagine, having a stronger link to their homeland, native homeland China, are more susceptible to communism than the native Thais.

Teacher 25T

However, there are some other large groups. Spread throughout China, through Thailand, are a large number of Chinese. In the capital city of Bangkok, about half the population is Chinese. That means approximately one million people there are Chinese. These Chinese have not really mixed well with the population, with the Thai population, and places they've been. They've tended to stay to themselves, as the overseas Chinese do in most Asian countries. They also have tended to work in certain areas. For example, in commerce, and to a great extent they dominate commerce, that is, they're the ones who are the businessmen, traders, the bankers, in Thailand. And in this sense, they probably represent the wealthiest ethnic group in Thailand. That is to say, in terms of money, they would be the upper crust rather than the Thai people.

Teacher 12T

There are some areas of unrest even though the picture looks rosy so far. The Chinese tend to dominate the commerce of the country. In Bangkok itself, over half the population is Chinese, and they take care of most of the commercial practices. And this is a bad thing in relation to the future of Thailand because the Thai people themselves can't produce leaders and they have to, you might say, import them. The influence of the Chinese, because of their commercial wealth, is felt in politics, and this of course then is another area of tenderness because the Chinese seem to be more prone to the communist way of thinking. They seem to be more militant than the easy-going Thai who are, because of their religion, kinda passive and easy-going.

Teacher 36T

Now the people here are Thai, but they have problems with other Asiastic peoples. One of these particular problems is with the Chinese. The Chinese in Bangkok in this area are the dominant force in the economy. They are in control of the economy, industries that they have, the shops, and so forth. The businessman in Thailand in this area, the central part, is Chinese. The Thai people are Buddhist, easy-going, sort of relaxed, not too ambitious. The Chinese are very ambitious and want to get ahead, so they dominate. Recently, when they had a company offer the Thai people 25 scholarships, and the scholarships were competitive, and of the 25 scholarships, 23 were won by the Chinese minority and two were won by the Thai people on a simple competitive examination. So this is one of their problems. The Thai people are easy-going and the Chinese are dominant, but the problems here with the Chinese is that they can be influenced by Communist China and it seems that the Chinese here are susceptible to influence from Communism, especially from their mother country, Communist China. So the Thais are upset by the fact that the Chinese are more this way. The Thais themselves because of their Buddhist faith and the fact of easy life found out that Communism is not too attractive to them, so they're not as willing to fall for the Communist line.